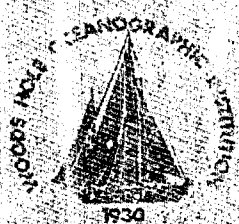WHOI-88-26

Woods Hole
Oceanographic
Institution

Multiple Source Location Estimation
Using the EM Algorithm

by

Ehud Weinstein and Meir Feder

July 1988

Technical Report

Multiple Source Location Estimation
Using the EM Algorithm

by

Ehud Weinstein and Meir Feder

Woods Hole Oceanographic Institution
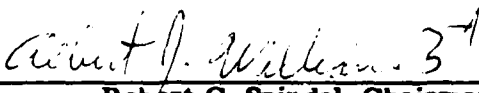Woods Hole, Massachusetts 02543

July 1986

Technical Report

Approved for Distribution:

_____
Robert C. Spindel, Chairman
Department of Ocean Engineering

Abstract

We present a computationally efficient scheme for multiple source location estimation based on the EM Algorithm. The proposed scheme is optimal in the sense that it converges iteratively to the exact Maximum Likelihood estimate of all the unknown parameters simultaneously. The method can be applied to a wide range of problems arising in signal and array processing.

I. Introduction

The Maximum Likelihood (ML) method is widely regarded as the optimal procedure in parameter estimation. However, when applied to problems involving composite signals and a large number of parameters, the ML method tends to be computationally complex and time consuming. As an example, consider the problem of multiple source location estimation by signal observations using an array of spatially distributed sensors. The received signals are composed of the contributions from the various signal sources, observed in the presence of additive noises. The ML estimate of the various source location parameters jointly requires the solution of a complicated, non-linear optimization problem in several unknowns. Problems of that nature frequently arise in time series analysis, array processing, radar/sonar, acoustical and geophysical signal processing.

In this report we develop a computationally efficient scheme for parameter estimation of composite signals, based on the EM algorithm. The proposed scheme is optimal in the sense that it converges iteratively to the exact ML estimate of all the unknown parameters simultaneously.

II. Maximum Likelihood Estimation and the EM Algorithm

Let $\underline{Y}$ denote the data vector possessing the probability density $f_{\underline{Y}}(\underline{y}; \underline{\theta})$ indexed by the parameter vector $\underline{\theta} \epsilon \Theta$. Θ is a subset of the Euclidean K-space. Given an observed $\underline{y}$, the ML estimate $\hat{\underline{\theta}}_{ML}$ is the value of $\underline{\theta}$ that maximizes the log-likelihood, that is

$$\underset{\underline{\theta} \epsilon \Theta}{Max} \; \log f_{\underline{Y}}(\underline{y}; \underline{\theta}) \longrightarrow \hat{\underline{\theta}}_{ML} \qquad (1)$$

Suppose the data vector $\underline{Y}$ can be viewed as being incomplete, and we can specify some "complete" data $\underline{X}$ related to $\underline{Y}$ by :

$$H(\underline{X}) = \underline{Y} \qquad (2)$$

where $H(\cdot)$ is a non-invertable (many-to-one) transformation. In the multiple source location problem, the 'complete' data could be the observation of the various source signals <u>separately</u>, where the observed (incomplete) data is the <u>sum</u> of the signal contributions from the various sources.

The EM algorithm is directed at finding the solution to (1); however, it does so by making an essential use of the complete data specification. The algorithm is basically an iterative method. It starts with an initial guess $\underline{\theta}^{(0)}$, and let $\underline{\theta}^{(n+1)}$ be defined inductively by :

$$\underset{\underline{\theta}}{Max} \; E\left\{ \log f_{\underline{X}}(\underline{x}; \underline{\theta}) / \underline{Y} = \underline{y} ; \underline{\theta}^{(n)} \right\} \longrightarrow \underline{\theta}^{(n+1)} \qquad (3)$$

where $f_{\underline{X}}(\underline{x}; \underline{\theta})$ is the probability density of $\underline{X}$, and $E\{\cdot / \underline{Y}=\underline{y}; \underline{\theta}^{(n)}\}$ denotes the conditional expectation given $\underline{y}$, computed using the parameter value $\underline{\theta}^{(n)}$. The heuristic idea here is that we would like to choose $\underline{\theta}$ that maximizes $\log f_{\underline{X}}(\underline{x}; \underline{\theta})$, the log-likelihood of the complete data. However, since $\log f_{\underline{X}}(\underline{x}; \underline{\theta})$ is not available to us (because the complete data is not available), we maximize instead its expectation, given the observed data $\underline{y}$. Since we have used the current estimate $\underline{\theta}^{(n)}$ rather than the actual value of $\underline{\theta}$ which is unknown, the conditional expectation is not exact. Thus, tne algorithm iterates, using each new parameter estimate to improve the conditional expectation on the next iteration cycle and thus to improve the next parameter estimate.

The EM algorithm was first presented in [1], where it is shown that under the usual regularity conditions, the algorithm converges to the desired result, that is $\underline{\theta}^{(n)} \longrightarrow \hat{\underline{\theta}}_{ML}$, where each iteration increases the likelihood. The basic considerations leading to the EM algorithm (Eq.(3)) are outlined in Appendix A for the convenience of the reader. The rate of convergence of the algorithm is exponential, depending on the fraction of the covariance of the "complete" data that can be predicted using the observed data. If that fraction is small, the rate of convergence tends to be slow, in which case one could use standard numerical methods to accelerate the algorithm.

We note that the EM algorithm is not uniquely defined. The transformation $H(\cdot)$ relating $\underline{X}$ to $\underline{Y}$ can be any non-invertable transformation. Obviously, there are many possible "complete" data specifications that will generate the observed data. Thus, the EM algorithm can be implemented in many possible ways. The final outcome, which is the ML estimate, is completely

unaffected by the way in which $H(\cdot)$ is specified (i.e., the choice of "complete" data). However, the choice of $H(\cdot)$ may critically affect the complexity and rate of convergence of the algorithm, and the unfortunate choice of $H(\cdot)$ may yield a completely useless algorithm.

We shall proceed as follows: First we develop the EM algorithm for the Linear-Gaussian case. This case covers a wide range of applications. Then we show that for the class of problems of interest here, there is a natural choice of the "complete" data, leading to a surprisingly simple algorithm to extract the ML estimates.

III. The Linear-Gaussian Case

Suppose that $\underline{Y} = H\underline{X}$, where H is a $m \times n$ matrix ($m < n$), and $\underline{X}$ possesses the following multivariate Gaussian probability density:

$$f_{\underline{X}}(\underline{x};\underline{\theta}) = \left[\det\left(\frac{2\pi}{\lambda}\Lambda(\underline{\theta}) \right) \right]^{-\frac{\lambda}{2}} \exp\left[-\frac{\lambda}{2}(\underline{x}-\underline{m}(\underline{\theta}))^{*}\Lambda^{-1}(\underline{\theta})(\underline{x}-\underline{m}(\underline{\theta})) \right] \tag{4}$$

where $\lambda = 1$ if $\underline{X}$ is real-valued, $\lambda = 2$ if $\underline{X}$ is complex valued, and the superscript $*$ denotes the conjugate-transpose operation. We shall refer to this case as the Linear-Gaussian case. Taking the logarithm or (4), we obtain:

$$\log f_{\underline{X}}(\underline{x};\underline{\theta}) = -\frac{\lambda}{2}\log\det\left(\frac{2\pi}{\lambda}\Lambda(\underline{\theta}) \right) - \frac{\lambda}{2}(\underline{x}-\underline{m}(\underline{\theta}))^{*}\Lambda^{-1}(\underline{\theta})(\underline{x}-\underline{m}(\underline{\theta}))$$

$$= -\frac{\lambda}{2}\log\det\left(\frac{2\pi}{\lambda}\Lambda(\underline{\theta}) \right) - \frac{\lambda}{2}\underline{m}^{*}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta})$$

$$+ \frac{\lambda}{2}\underline{x}^{*}\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta}) + \frac{\lambda}{2}\underline{m}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{x} - \frac{\lambda}{2}tr\left(\Lambda^{-1}(\underline{\theta})\underline{x}\,\underline{x}^{*} \right) \tag{5}$$

where $tr(\cdot)$ stands for the trace of the bracketed matrix. Thus,

$$E\left\{ \log f_{\underline{X}}(\underline{x};\underline{\theta}) / \underline{Y}=\underline{y}; \underline{\theta}^{(n)} \right\} = -\frac{\lambda}{2}\log\det\left(\frac{2\pi}{\lambda}\Lambda(\underline{\theta}) \right) - \frac{\lambda}{2}\underline{m}^{*}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta})$$

$$+ \frac{\lambda}{2}\underline{x}^{(n)*}\Lambda^{-1}(\underline{\theta})\underline{m}(\underline{\theta}) + \frac{\lambda}{2}\underline{m}^{*}(\underline{\theta})\Lambda^{-1}(\underline{\theta})\underline{x}^{(n)} - \frac{\lambda}{2}tr\left(\Lambda^{-1}(\underline{\theta})\Psi^{(n)} \right) \tag{6}$$

where $\underline{X}^{(n)} \triangleq E\{\underline{X}/\underline{Y}=\underline{y}; \underline{\theta}^{(n)}\}$ and $\underline{\Psi}^{(n)} \triangleq E\{\underline{X}\underline{X}^{*}/\underline{Y}=\underline{y}; \underline{\theta}^{(n)}\}$.

Since $\underline{X}$ and $\underline{Y}$ are jointly Gaussian, these conditional expectations are readily available in the literature (e.g. [2] , Chap. 5):

$$\underline{X}^{(n)} = \underline{m}(\underline{\theta}^{(n)}) + \Gamma(\underline{\theta}^{(n)})[\underline{y} - H \cdot \underline{m}(\underline{\theta}^{(n)})] \tag{7}$$

$$\underline{\Psi}^{(n)} = [I - \Gamma(\underline{\theta}^{(n)}) \cdot H] \Lambda(\underline{\theta}^{(n)}) + \underline{X}^{(n)} \underline{X}^{(n)*} \tag{8}$$

where I is the identity matrix, and $\Gamma(\underline{\theta})$ is the "Kalman Gain" defined by:

$$\Gamma(\underline{\theta}) = \Lambda(\underline{\theta}) H^{*}[H\Lambda(\underline{\theta})H^{*}]^{-1} \tag{9}$$

The EM algorithm is completely specified by Eqs. (6) - (9); the algorithm iterates between calculating $\underline{X}^{(n)}$ and $\underline{\Psi}^{(n)}$ and maximizing the expression in (6) with respect to $\underline{\theta}$, where each iteration increases the likelihood. We observe that $E\{log f_{\underline{X}}(\underline{x};\underline{\theta})/\underline{Y}=\underline{y}; \underline{\theta}^{(n)}\}$ (Eq. (6)) and $log f_{\underline{X}}(\underline{x};\underline{\theta})$ (Eq. (5)) have the same dependence on $\underline{\theta}$. Maximizing $E\{log f_{\underline{X}}(\underline{x};\underline{\theta})/\underline{Y}=\underline{y}; \underline{\theta}^{(n)}\}$ with respect to $\underline{\theta}$ is the same as maximizing $log f_{\underline{X}}(\underline{x};\underline{\theta})$ with respect to $\underline{\theta}$. Hence, the EM algorithm essentially requires the ML solution in the $\underline{X}$ model which might be significantly simpler than the direct ML solution in the $\underline{Y}$ model.

IV. Application to Signal Processing

Let the mathematical model characterizing the observed signal(s) be given by

$$\underline{y}(t) = \underline{s}(t; \underline{\theta}) + \underline{\eta}(t) \qquad T_i \leq t \leq T_f \qquad (10.1)$$

where $\underline{s}(t; \underline{\theta})$ are the vector signals received in the absence of noise. We shall be concerned with $\underline{s}(t; \underline{\theta})$ composed of

$$\underline{s}(t; \underline{\theta}) = \sum_{k=1}^{K} \underline{s}_k(t; \underline{\theta}_k) \qquad (10.2)$$

where $\underline{\theta}_k$ is the parameter vector associated with the k signal component. The specific problem we have in mind is multiple source location estimation, in which case the $\underline{s}_k(t; \underline{\theta}_k)$ are the array signals received from the k source, and $\underline{\theta}_k$ are the corresponding source location parameters (bearing and range, velocity components, etc.) In addition, the model given by (10) covers a wide range of problems arising in array and signal processing.

We shall now consider the joint estimation of the various $\underline{\theta}_k$ for the case of deterministic signals and for the case of stochastic Gaussian signals separately.

A. Deterministic Signals

Consider the model of (10) under the following assumptions:

- The $\underline{s}_k(t; \underline{\theta}_k)$ are conditionally known to the observer (i.e. , given $\underline{\theta}_k$, we can construct $\underline{s}_k(t; \underline{\theta}_k)$).

- $\underline{\eta}(t)$ are vector zero-mean white Gaussian stochastic processes whose covariance matrix is $E\{\underline{\eta}(t)\underline{\eta}^{\dagger}(\Delta)\} = Q \cdot \delta(t-\Delta)$, where Q is a constant matrix and $\delta(\tau)$ is the impulse function.

Under the above assumptions, the log-likelihood function is given by

$$\log f_y(\underline{y};\theta) = c - \frac{\lambda}{2}\int_{T_i}^{T_f}[\underline{y}(t)-\sum_{k=1}^{K}\underline{s}_k(t;\underline{\theta}_k)]^*\overset{-1}{Q}[\underline{y}(t)-\sum_{k=1}^{K}\underline{s}_k(t;\underline{\theta}_k)]dt \qquad (11)$$

where c is a normalizing constant. The result in (11) is a straightforward

multi-channel extension of the known (deterministic) signal in a white Gaussian

noise problem ([3], Chap. 4). In case we are given only a discrete set of

observations $\underline{y}(t_i)$ $i=1, 2, \dots N$, the log-likelihood is still given by

(11), where the integral over t is replaced by the sum over the t_i's .

Thus, the joint ML estimation of the various $\underline{\theta}_k$'s is obtained as the solution to:

$$\underset{\underline{\theta}_1, \underline{\theta}_2, \dots \underline{\theta}_k}{Min} \left\{ \int_{T_i}^{T_f}[\underline{y}(t)-\sum_{k=1}^{K}\underline{s}_k(t;\underline{\theta}_k)]^*\overset{-1}{Q}[\underline{y}(t)-\sum_{k=1}^{K}\underline{s}_k(t;\underline{\theta}_k)]dt \right\}$$

$$\Longrightarrow \hat{\underline{\theta}}_1, \hat{\underline{\theta}}_2, \dots \hat{\underline{\theta}}_K \qquad (12)$$

In the case of discrete observations, the corresponding optimization problem is:

$$\underset{\underline{\theta}_1, \underline{\theta}_2, \dots \underline{\theta}_K}{Min} \left\{ \sum_{i=1}^{N}[\underline{y}(t_i)-\sum_{k=1}^{K}\underline{s}_k(t_i;\underline{\theta}_k)]^*\overset{-1}{Q}[\underline{y}(t_i)-\sum_{k=1}^{K}\underline{s}_k(t_i;\underline{\theta}_k)]dt \right\}$$

$$\Longrightarrow \hat{\underline{\theta}}_1, \hat{\underline{\theta}}_2, \dots \hat{\underline{\theta}}_K \qquad (13)$$

In either case, the ML method calls for the optimization of the log-likelihood

function with respect to all the $\underline{\theta}_k$'s jointly. Of course, brute force can

always be used to solve the problem, evaluating the objective function on a coarse

grid to roughly locate the global minimum, and then applying the Gauss method or

Newton-Raphson or some other iterative gradient-search algorithm. However, when

applied to the problem at hand, these methods tend to be very complex and

computationally time consuming.

Having the EM algorithm in mind, we would like to simplify the optimization problem associated with the direct ML approach. To apply the algorithm to the problem at hand, the first step is to specify the "complete" data. A natural choice of the "complete" data is obtained by decomposing $\underline{y}(t)$ into

$$\underline{y}(t) = \sum_{k=1}^{K} \underline{x}_k(t) \tag{14.1}$$

where

$$\underline{x}_k(t) = \underline{s}_k(t; \underline{\theta}_k) + \underline{n}_k(t) \tag{14.2}$$

and the $\underline{n}_k(t)$ are chosen to be mutually uncorrelated, zero-mean Gaussian vector stochastic processes satisfying

$$\sum_{k=1}^{K} \underline{n}_k(t) = \underline{n}(t) \tag{14.3}$$

Let the complete data $\underline{x}(t)$ be composed of the various $\underline{x}_k(t)$'s . Then from (14.1), the relation between the "complete" data and the observed (incomplete) data is given by

$$\underline{y}(t) = H \cdot \underline{x}(t) \tag{15.1}$$

where

$$H = [\underbrace{I \vdots I \vdots \ldots \vdots I}_{K \text{ terms}}] \tag{15.2}$$

Since the "complete" data is Gaussian, and the relation between $\underline{X}(t)$ and $\underline{y}(t)$ is linear, the results developed for the Linear-Gaussian case can be applied here. The detailed derivation is given in Appendix B. The resulting algorithm is:

<u>E-step</u>   For $k = 1, 2, \ldots K$, compute:

$$\underline{X}_k^{(n)}(t) = \underline{g}_k(t; \underline{\theta}^{(m)}) + \beta_k \left[\underline{y}(t) - \sum_{b=1}^{K} \underline{g}_k(t; \underline{\theta}_k^{(m)}) \right] \tag{16.1}$$

<u>M-step</u>   For $k = 1, 2, \ldots K$

$$\underset{\underline{\theta}_k}{Min} \int_{T_i}^{T_f} \left[\underline{X}_k^{(m)}(t) - \underline{g}_k(t; \underline{\theta}_k) \right]^* \underline{Q}^{-1} \left[\underline{X}_k^{(m)}(t) - \underline{g}_k(t; \underline{\theta}_k) \right] dt \longrightarrow \underline{\theta}_k^{(n+1)} \tag{16.2}$$

where the $\beta_k's$ are real-valued positive scalars satisfying

$$\sum_{b=1}^{K} \beta_k = 1 \tag{16.3}$$

In the discrete data case, the algorithm takes the form:

<u>E-step</u>   Compute:

$$\underline{X}_k^{(m)}(t_i; \underline{\theta}_k) = \underline{g}_k(t_i; \underline{\theta}_k^{(m)}) + \beta_k \left[\underline{y}(t_i) - \sum_{b=1}^{K} \underline{g}_k(t_i; \underline{\theta}_k^{(m)}) \right] \tag{17.1}$$

<u>M-step</u>

$$\underset{\underline{\theta}_k}{Min} \sum_{i=1}^{N} \left[\underline{X}_k^{(m)}(t_i; \underline{\theta}_k) - \underline{g}_k(t_i; \underline{\theta}_k) \right]^* \underline{Q}^{-1} \left[\underline{X}_k^{(m)}(t_i) - \underline{g}_k(t_i; \underline{\theta}_k) \right] \longrightarrow \underline{\theta}_k^{(n+1)} \tag{17.2}$$

The optimization required by (16.2) is, in fact, the optimization problem associated with the ML estimation of the $\underline{\theta}_k$ given <u>separate</u> observations of the $\underline{g}_k(t; \underline{\theta}_k)$.   An entirely parallel statement can be made with respect

to the optimization required by (17.2). Thus, the algorithm decouples the optimization associated with the direct ML approach into the K separate ML optimizations as illustrated in Fig. 1. Hence, the complexity of the algorithm is completely unaffected by the assumed number of signal components. As K increases, we have to increase the number of ML processors in parallel; however, each processor is maximized separately. Since the algorithm is based on the EM method, it must converge to the <u>exact</u> ML estimate of the various θ_k's simultaneously, where each iteration increases the likelihood.

We note that the β_k's must satisfy the constraint (16.3), but otherwise they are arbitrary free variables in the algorithm. The choice of the β_k's does not affect the value of the final estimates; however, they can be used to control the rate of convergence of the algorithm.

Gaussian Signals

Consider the model of (10) under the following assumptions:

- The $\underline{s}_k(t; \underline{\theta}_k)$ $k = 1, 2, \ldots K$ are mutually uncorrelated, wide sense stationary (WSS), zero-mean Gaussian vector stochastic processes with the spectral density matrices $S_k(\omega; \underline{\theta}_k)$ $k = 1, 2, \ldots K$ respectively.

- $\underline{n}(t)$ is uncorrelated, WSS, zero-mean and Gaussian with the spectral density matrix $N(\omega)$.

If the observation interval $T = T_f - T_i$ is long compared with the correlation time (inverse bandwidth) of the signals and the noises, the log-likelihood is given by (see Appendix C)

$$\log f_{\underline{y}}(\underline{y}; \underline{\theta}) = -\sum_{\ell} \left[\log \det P(\omega_\ell; \underline{\theta}) + Y^*(\omega_\ell) P^{-1}(\omega_\ell; \underline{\theta}) Y(\omega_\ell) \right] \qquad (18.1)$$

where

$$\underline{Y}(\omega_\ell) = \frac{1}{\sqrt{T}} \int_{T_i}^{T_f} \underline{y}(t) e^{-j\omega_\ell t} dt \qquad \omega_\ell = \frac{2\pi}{T} \cdot \ell \qquad (18.2)$$

and

$$P(\omega; \underline{\theta}) = \sum_{k=1}^{K} S_k(\omega; \underline{\theta}_k) + N(\omega) \qquad (18.3)$$

The summation in (18.1) is carried over all ω_ℓ in the signal frequency band.

In the case of discrete observations $\underline{y}_i = \underline{y}(i \cdot \Delta t)$, the log-likelihood is still given by (18.1), where $\{\underline{Y}(\omega_\ell)\}_{\ell=0}^{N-1}$ $\omega_\ell = 2\pi\ell/N$ are the discrete Fourier transform (DFT) of $\{\underline{y}_i\}_{i=0}^{N-1}$, and $P(\omega; \underline{\theta})$ is given by (18.3), where $S_k(\omega; \underline{\theta}_k)$ is the spectral matrix of the corresponding sampled signal, and $N(\omega)$ is the spectral matrix of the sampled noise. Thus, in either case, the joint ML estimation of the various $\underline{\theta}_k$'s is obtained as the solution to:

$$\underset{\underline{\theta}_1, \underline{\theta}_2, \dots \underline{\theta}_K}{\mathrm{Min}} \left\{ \sum_\ell \left[\log \det P(\omega_\ell; \underline{\theta}) + \underline{Y}^*(\omega_\ell) P^{-1}(\omega_\ell; \underline{\theta}) \underline{Y}(\omega_\ell) \right] \right\}$$
$$\Longrightarrow \hat{\underline{\theta}}_1, \hat{\underline{\theta}}_2, \dots \hat{\underline{\theta}}_K \qquad (19)$$

Thus, as in the deterministic signal case, the direct ML approach requires the joint optimization of the objective function with respect to all the $\underline{\theta}_k$'s. Using the EM method, we would like to simplify the required optimization. Using the same definition for the "complete" data as in the deterministic signal case (Eq.(14)) and applying the results developed for the Linear-Gaussian case, we obtain the following algorithm:

E-step   Compute:

$$\Psi_k^{(n)}(\omega_\ell) = B_k(\omega_\ell; \underline{\theta}_k^{(m)}) \left[I - \overset{-1}{P}(\omega_\ell; \underline{\theta}^{(m)}) B_k(\omega_\ell; \underline{\theta}_k^{(m)}) \right] +$$

$$+ B_k(\omega_\ell; \underline{\theta}_k^{(m)}) \overset{-1}{P}(\omega_\ell; \underline{\theta}^{(m)}) \underline{Y}(\omega_\ell) \underline{Y}^*(\omega_\ell) \overset{-1}{P}(\omega_\ell; \underline{\theta}^{(m)}) B_k(\omega_\ell; \underline{\theta}_k^{(m)}) \quad (20.1)$$

M-step

$$\underset{\underline{\theta}_k}{Min} \sum_\ell \left[\log \det B_k(\omega_\ell; \underline{\theta}_k) + tr \left(B_k(\omega; \underline{\theta}_k) \Psi_k^{(n)}(\omega_\ell) \right) \right] \Longrightarrow \underline{\theta}_k^{(m+1)} \quad (20.2)$$

where we define

$$B_k(\omega; \underline{\theta}_k) = S_k(\omega; \underline{\theta}_k) + \beta_k \cdot N(\omega) \quad (20.3)$$

and the β_k's are arbitrary real-valued scalars satisfying the condition (16.3). The detailed considerations leading to Eq.(20) are given in Appendix D. Perhaps the most striking feature of the algorithm is that it decouples the complex optimization associated with the direct ML approach into the K separate ML optimizations as illustrated in Fig. 1. Thus, as in the deterministic signal case, we obtain a considerable simplification in estimator structure and computations.

Example: Multiple Source Location Estimation

The basic system of interest here consists of several spatially distributed signal sources and an array of several spatially distributed sensors. The observed signals can be modelled using (10), where $\underline{\theta}_k(t; \underline{\theta}_k)$ are the array signals received

from the k source, and $\underline{\rho}_k$ are the source location parameters (bearing and range, velocity components, etc.). Using the proposed scheme (Fig. 1), we obtain the exact ML estimates of the various source location parameters, while circumventing the complex multiple parameter optimization associated with the direct ML approach. For more details we refer the interested reader to [4] and [5]. To simplify the example, we have considered the following situation:

- The array consists of 10 co-linear and evenly spaced sensors. The observed data consists of 100 time samples at each sensor output.

- There are three far-field sources at bearings 0.,0.15 and 0.3 radians, measured relative to the Boresight direction. Each source radiates a triangular pulse whose power (energy per sample) is normalized to 1, and whose duration is 20 time points.

- The additive noises are uncorrelated from sensor to sensor with the same spectral level of 0.5 (that is, the post-integration SNR per channel is approximately 16 db).

We have also assumed that $\Delta L / u \cdot \Delta t = 10$, where Δt is the sampling period, ΔL is the spacing between adjacent sensors, and u is the velocity of propagation in the medium. The problem is to estimate the various source bearings simultaneously.

In Fig. 2 we have plotted the conventional beam-former output as a function of bearing. As we can see, the standard method cannot resolve the various sources. In Fig. 3 we have plotted the outcome using our algorithm. We clearly see that after about 5 iterations, the algorithm essentially converges to the true bearings, and the three sources are resolved correctly.

V. Conclusions

We have presented a computationally efficient scheme for parameter estimation of composite signals based on the EM algorithm. The proposed scheme is optimal in the sense that it converges iteratively to the <u>exact</u> <u>Maximum</u> <u>Likelihood</u> estimate of all the unknown parameters simultaneously. The method can be applied to a wide range of problems arising in signal and array processing.

Appendix A   The EM Algorithm

Let $H(\underline{X}) = \underline{Y}$ where $H(\cdot)$ is a non-invertable (many-to-one) transformation. Express densities

$$f_{\underline{X}}(\underline{x}; \underline{\theta}) = f_{\underline{Y}}(H(\underline{x}); \underline{\theta}) \cdot f_{\underline{X}/\underline{Y}=\underline{y}}(\underline{x}; \underline{\theta}) \qquad (A.1)$$

Taking the logarithm on both sides of (A.1), we obtain

$$\log f_{\underline{Y}}(H(\underline{x}); \underline{\theta}) = \log f_{\underline{X}}(\underline{x}; \underline{\theta}) - \log f_{\underline{X}/\underline{Y}=\underline{y}}(\underline{x}; \underline{\theta}) \qquad (A.2)$$

Denote $\underline{y} = H(\underline{x})$ and taking conditional expectations given $\underline{Y} = \underline{y}$ for a parameter value $\underline{\theta}'$, we obtain

$$\log f_{\underline{Y}}(\underline{y}; \underline{\theta}) = E\{\log f_{\underline{X}}(\underline{x}; \underline{\theta}) / \underline{Y} = \underline{y}; \underline{\theta}'\} - $$
$$- E\{\log f_{\underline{X}/\underline{Y}=\underline{y}}(\underline{x}; \underline{\theta}) / \underline{Y}=\underline{y}; \underline{\theta}'\} = U(\underline{\theta}, \underline{\theta}') - V(\underline{\theta}, \underline{\theta}') \qquad (A.3)$$

Invoking the Jensen's inequality, we have that $V(\underline{\theta}, \underline{\theta}') \leq V(\underline{\theta}', \underline{\theta}')$. Hence

$$U(\underline{\theta}, \underline{\theta}') > U(\underline{\theta}', \underline{\theta}') \Longrightarrow \log f_{\underline{Y}}(\underline{y}; \underline{\theta}) > \log f_{\underline{Y}}(\underline{y}; \underline{\theta}') \qquad (A.4)$$

Eq.(A.4) forms the basis to the EM algorithm. The algorithm starts with an initial guess $\underline{\theta}^{(0)}$, and let $\underline{\theta}^{(n+1)}$ be defined inductively by

$$\underset{\underline{\theta}}{Max}\ U(\underline{\theta}, \underline{\theta}^{(n)}) \Longrightarrow \underline{\theta}^{(n+1)} \qquad (A.5)$$

Eqs. (A.5) and (3) are identical. We note that since $\underline{\theta}^{(n+1)}$ is the value of $\underline{\theta}$ that maximizes $U(\underline{\theta}, \underline{\theta}^{(n)})$, then according to (A.4), each iteration of the algorithm increases the value of $\log f_{\underline{Y}}(\underline{y}; \underline{\theta})$.

Appendix B.   Derivation of Eq.(16)

The "complete" data $\underline{X}(t)$ is given by

$$\underline{X}^*(t) = \left(\underline{X}_1^*(t)\,; \; \underline{X}_2^*(t)\,; \; \dots \; ; \; \underline{X}_K^*(t) \right) \tag{B.1}$$

where the $\underline{X}_k(t)$ are given by (14.2), rewritten here for reference

$$\underline{X}_k(t) = \underline{s}_k(t;\,\theta_k) + \underline{n}_k(t) \tag{B.2}$$

The $\underline{n}_k(t)$ are chosen to be statistically independent, zero-mean and Gaussian with the covariance $E\left\{ \underline{n}_k(t)\,\underline{n}_k^*(s) \right\} = Q_k \cdot \Gamma(t-s)$, where $Q_k = \beta_k \cdot Q$. The $\underline{X}_k(t)$ are, therefore, mutually independent and hence, following the usual considerstions ([3], Chap. 4), we obtain

$$\log f_{\underline{X}}(\underline{x};\,\theta) = \sum_{k=1}^{K} \log f_{\underline{X}_k}(\underline{x}_k;\,\theta_k)$$

$$= c - \sum_{k=1}^{K} \int_{T_i}^{T_f} [\underline{X}_k(t) - \underline{s}_k(t;\,\theta_k)]^* Q_k^{-1} [\underline{X}_k(t) - \underline{s}_k(t;\,\theta_k)]\,dt \tag{B.3}$$

Thus,

$$E\left\{ \log f_{\underline{X}}(\underline{x};\,\theta)\,/\,\underline{Y} = \underline{y}\,;\,\theta^{(n)} \right\} =$$

$$= c' - \sum_{k=1}^{K} \int_{T_i}^{T_f} [\underline{X}_k^{(n)}(t) - \underline{s}_k(t;\,\theta_k)]^* Q_k^{-1} [\underline{X}_k^{(n)}(t) - \underline{s}_k(t;\,\theta_k)]\,dt \tag{B.4}$$

where c' contains all the terms that are independent of $\underline{\theta}$. The $\underline{X}_k^{(n)}(t)$ are the components of $\underline{X}^{(n)}(t)$, to be computed from

$$\underline{X}^{(n)}(t) = E\left\{\underline{X}(t) \,/\, H\cdot\underline{X}(t) = \underline{y}(t)\,;\, \underline{\theta}^{(n)}\right\} \tag{B.5}$$

Using Eq.(7) in (B.4) and following some straightforward matrix manipulations, we obtain (16.1). Now, since $\underline{\theta}_k$ enters the right side of (B.4) only through the k term in the sum, the joint maximization of (B.3) with respect to the various $\underline{\theta}_k$'s decouples into the K separate optimizations as suggested by (16.2).

Appendix C.   Derivation of Eq.(18)

Let the data vector $\underline{Y}$ be denerated by Fourier analyzing $\underline{y}(t)$, that is

$$\underline{Y}^* = (\ \underline{Y}^*(\omega_1) \ \vdots \ \underline{Y}^*(\omega_2) \ \vdots \ \dots \qquad\qquad) \qquad\qquad (C.1)$$

where the $\underline{Y}(\omega_\ell)$ are defined by (18.2). Since $\underline{y}(t)$ is Gaussian and the $\underline{Y}(\omega_\ell)$ are obtained by performing linear operations on $\underline{y}(t)$, then the $\underline{Y}(\omega_\ell)$ are jointly Gaussian. Since both signals and noises are assumed to be WSS, then for observation interval T long compared with the correlation time of the signals and the noises, the $\underline{Y}(\omega_\ell)$ are mutually uncorrelated (and thus, in the Gaussian case, statistically independent) with the covariance matrix given by

$$E\left\{ \underline{Y}(\omega_\ell) \underline{Y}^*(\omega_\ell) \right\} = \underline{P}(\omega_\ell; \underline{\theta}) \qquad\qquad (C.2)$$

where $\underline{P}(\omega_\ell;\underline{\theta})$ is defined by (18.2). It follows that

$$\ell_{\underline{Y}}(\underline{y};\underline{\theta}) = \prod_\ell \frac{1}{\det \underline{P}(\omega_\ell;\underline{\theta})} e^{-\underline{Y}^*(\omega_\ell) \underline{P}^{-1}(\omega_\ell;\underline{\theta}) \underline{Y}(\omega_\ell)} \qquad\qquad (C.3)$$

Taking the logarithm on both sides of (C.3) immediately yields (18.1).

Appendix D.   Derivation of Eq.(20)

Let the "complete" data $\underline{X}$ be generated by Fourier analyzing the various $\underline{X}_k(t)$, that is

$$\underline{X} = \left(\underline{X}_1^*(\omega_1) \,\vert\, \underline{X}_2^*(\omega_1) \,\vert\, \ldots \,\vert\, \underline{X}_k^*(\omega_1) \,\vert\, \underline{X}_1^*(\omega_2) \,\vert\, \underline{X}_2^*(\omega_2) \,\vert\, \ldots \,\vert\, \underline{X}_k^*(\omega_2) \,\vert\, \ldots \right) \qquad (D.1)$$

where $\underbrace{\qquad\qquad}_{\underline{X}(\omega_1)}$ $\underbrace{\qquad\qquad}_{\underline{X}(\omega_2)}$

$$\underline{X}_k(\omega_\ell) = \frac{1}{\sqrt{T}} \int_{T_i}^{T_f} \underline{X}_k(t) e^{-j\omega_\ell t} dt \qquad (D.2)$$

and $\underline{X}_k(t)$ are given by (14.2), rewritten here for reference

$$\underline{X}_k(t) = \underline{J}_k(t; \underline{\theta}_k) + \underline{\eta}_k(t) \qquad (D.3)$$

The $\underline{\eta}_k(t)$ are chosen to be statistically independent, zero-mean and Gaussian with the spectral density matrix $N_k(\omega) = \beta_k \cdot N(\omega)$. Since the $\underline{J}_k(t; \underline{\theta}_k)$ are assumed to be statistically independent, then the $\underline{X}_k(t)$ are statistically independent, in which case $\underline{X}_k(\omega_\ell)$ and $\underline{X}_m(\omega_p)$ are statistically independent whenever $k \neq m$. Furthermore, for long observation intervals and WSS processes, the $\underline{X}_k(\omega_\ell)$ at different frequencies are statistically uncorrelated (and thus independent) with the covariance matrix

$$E\left\{ \underline{X}_k(\omega_\ell) \underline{X}_k^*(\omega_\ell) \right\} = B_k(\omega_\ell; \underline{\theta}_k) \qquad (D.4)$$

where $B_k(\omega; \underline{\theta}_k)$ is defined by (20.3). It follows that

$$f_{\underline{X}}(\underline{Y};\underline{\theta}) = \prod_{k=1}^{K} \prod_{\ell} \frac{1}{\det B_k(\omega_\ell;\theta_k)} e^{-\underline{X}_k^*(\omega_\ell)B_k^{-1}(\omega_\ell;\theta_k)\underline{X}_k(\omega_\ell)} \qquad (D.5)$$

Taking the logarithm on both sides of (D.5), we obtain

$$\log f_{\underline{X}}(\underline{Y};\underline{\theta}) = -\sum_{k=1}^{K} \sum_{\ell} [\log \det B_k(\omega_\ell;\theta_k) + \underline{X}_k^*(\omega_\ell)B_k^{-1}(\omega_\ell;\theta_k)\underline{X}_k(\omega_\ell)] \qquad (D.6)$$

Taking the conditional expectation given $\underline{Y} = \underline{y}$ for a parameter value $\underline{\theta}^{(n)}$, we obtain

$$E\{\log f_{\underline{X}}(\underline{x};\underline{\theta})/\underline{Y}=\underline{y};\underline{\theta}^{(n)}\} =$$

$$= \sum_{k=1}^{K} \sum_{\ell} [\log \det B_k(\omega_\ell;\theta_k) + tr(B_k^{-1}(\omega_\ell;\theta_k)\Psi_k^{(n)}(\omega_\ell))] \qquad (D.7)$$

where $\Psi_k^{(n)}(\omega_\ell)$ is the (k,k) block matrix of $\Psi^{(n)}(\omega_\ell)$, to be computed from

$$\Psi^{(n)}(\omega_\ell) = E\{\underline{X}(\omega_\ell)\underline{X}^*(\omega_\ell)/H\underline{X}(\omega_\ell)=\underline{Y}(\omega_\ell);\underline{\theta}^{(n)}\} \qquad (D.8)$$

Using Eq.(8) in (D.8) and following some straightforward matrix manimpulations, we obtain (20.1). Since the (k,ℓ) term of the double sum on the right side of (D.7) depends only on $\underline{\theta}_k$, then the joint maximization of (D.7) with respect to the various $\underline{\theta}_k$ decouples into the K separate optimizations as suggested by (20.2).

References

[1] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm", Ann. of the Royal Stat. Soc., pp. 1-38, Dec. 1977.

[2] B.D.O. Anderson and J.B. Moore, Optimal Filtering, Prentice-Hall, Inc. N.J. 1979.

[3] H.L. Van-Trees, "Detection, Estimation and Modulation Theory", Part I, Wiley, N.Y. 1968.

[4] M. Feder and E. Weinstein, "Optimal Multiple Source Location Estimation Via the EM Algorithm", Proc. of the 1985 Int. Conf. on Acoust., Speech & Signal Processing (ICASSP'85), Vol. 4, pp. 1762-5, March 1985.

[5] M. Feder and E. Weinstein, "Multipath and Multiple Source Array Processing Via the EM Algorithm", Proc. of the 1986 Int. Conf. on Acoust., Speech & Signal Processing (ICASSP'86).
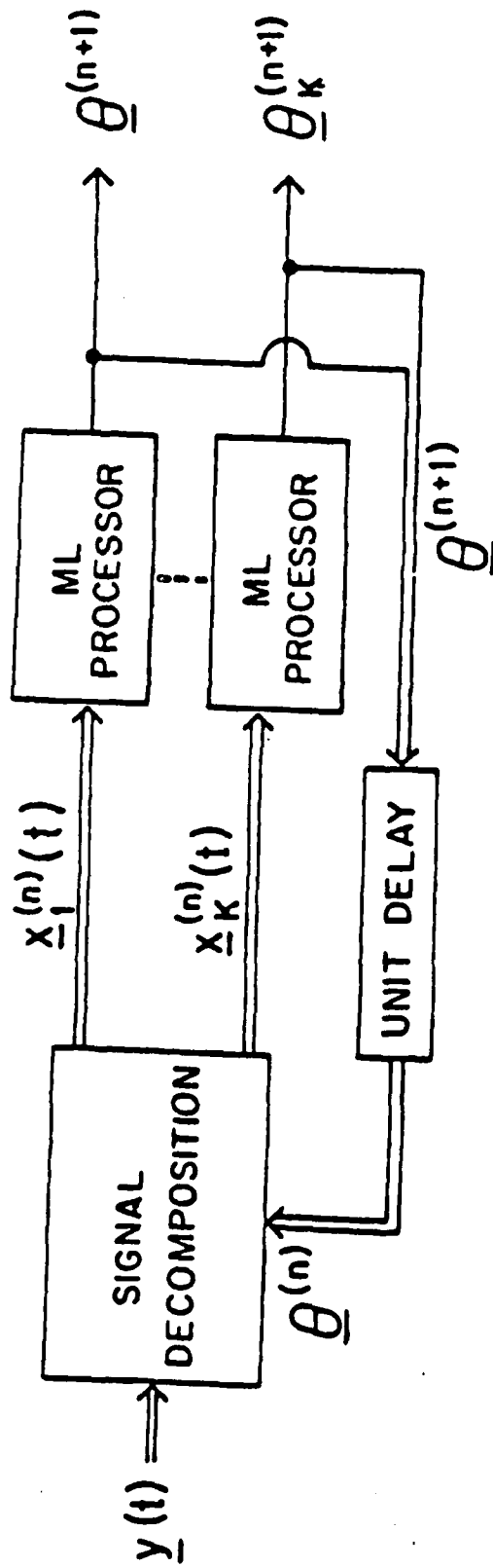
<u>Figure Captions</u>

Fig. 1:   Signal Processing Via the EM Algorithm.

Fig. 2:   Conventional Beamforming.

Fig. 3:   Multiple Source Location Estimation Via the Proposed Algorithm.

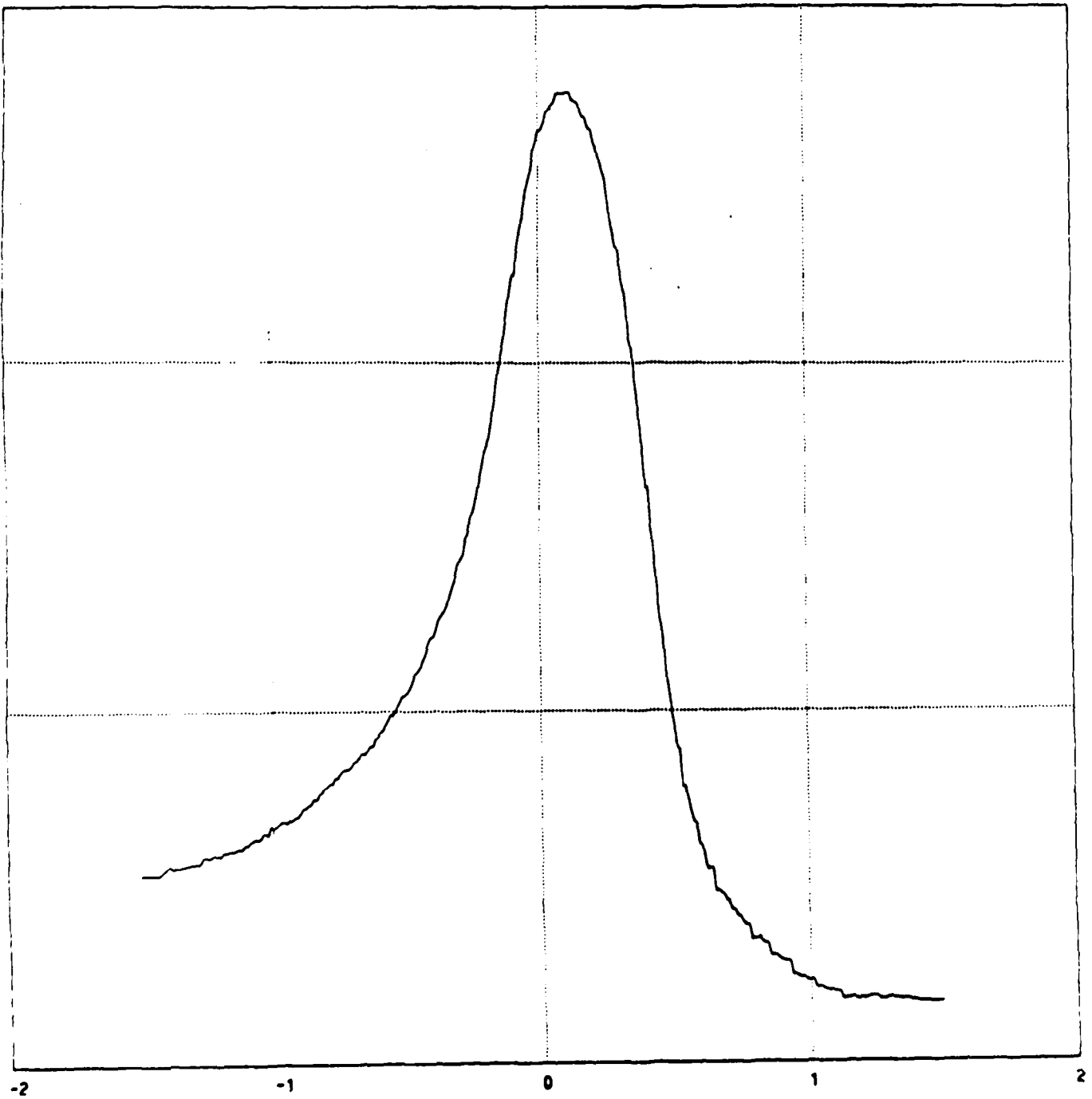Fig. 1: Signal Processing Via the EM Algorithm.
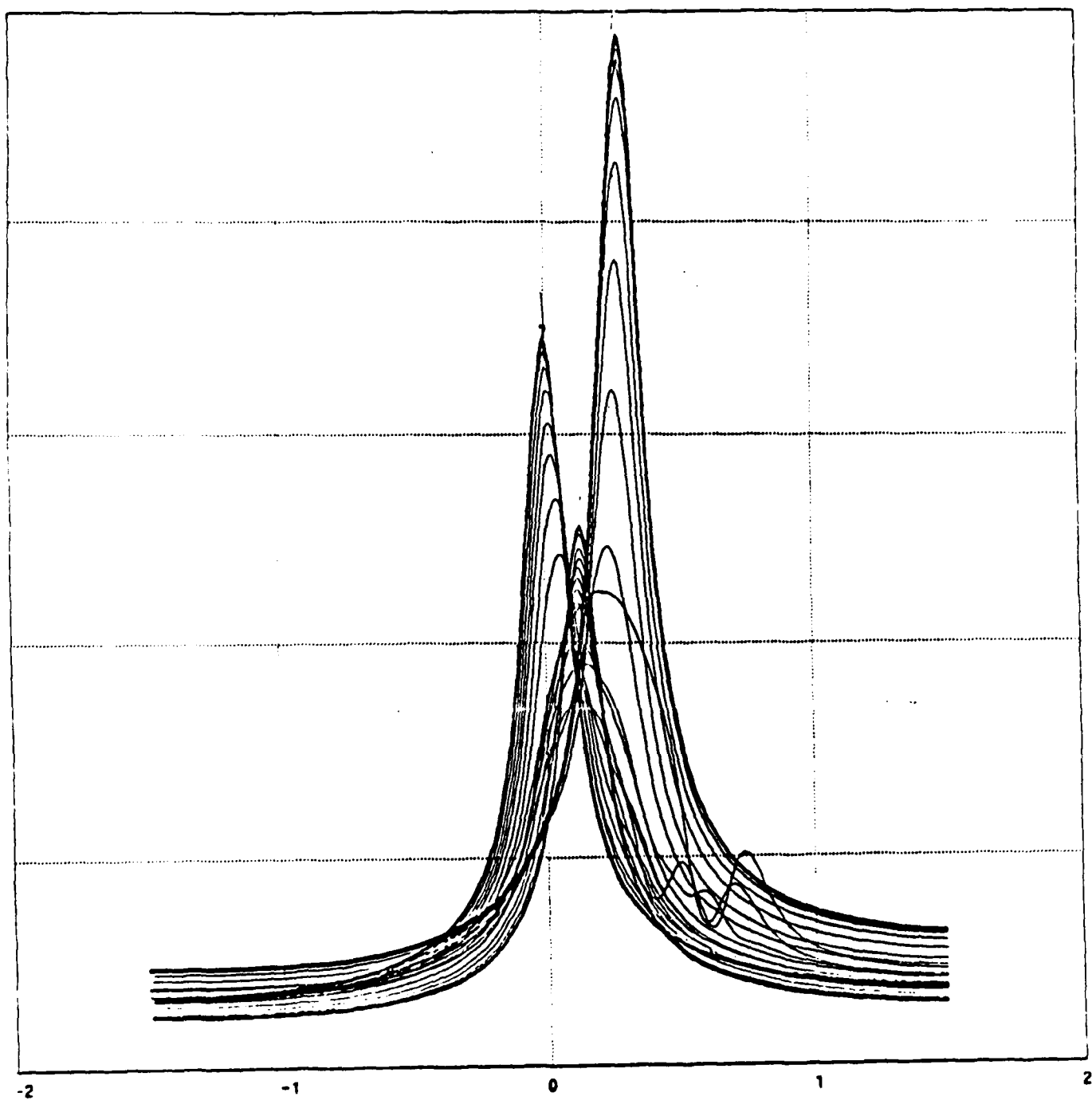
Fig. 2: Conventional Beamforming.

Fig. 3: Multiple Source Location Estimation Via the Proposed Algorithm.

DOCUMENT LIBRARY

April 9, 1985

DISTRIBUTION LIST FOR TECHNICAL REPORT EXCHANGE

Institute of Marine Sciences Library
University of Alaska
O'Neill Building
905 Koyukuk Ave., North
Fairbanks, AK

Attn: Stella Sanchez-Wade
Documents Section
Scripps Institution of Oceanography
Library, Mail Code C-075C
La Jolla, CA   92093

Hancock Library of Biology & Oceanography
Alan Hancock Laboratory
University of Southern California
University Park
Los Angeles, CA   90089-0371

Gifts & Exchanges
Library
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, NS, B2Y 4A2, CANADA

Office of the International
   Ice Patrol
c/o Coast Guard R & D Center
Avery Point
Groton, CT   06340

Library
Physical Oceanographic Laboratory
Nova University
8000 N. Ocean Drive
Dania, FL   33304

NOAA/EDIS Miami Library Center
4301 Rickenbacker Causeway
Miami, FL   33149

Library
Skidaway Institute of Oceanography
P.O. Box 13687
Savannah, GA   31416

Institute of Geophysics
University of Hawaii
Library Room 252
2525 Correa Road
Honolulu, HI   96822

Library
Chesapeake Bay Institute
4800 Atwell Road
Shady Side, MD   20876

MIT Libraries
Serial Journal Room 14E-210
Cambridge, MA   02139

Director, Ralph M. Parsons Laboratory
Room 48-311
MIT
Cambridge, MA   02139

Marine Resources Information Center
Bldg. E38-320
MIT
Cambridge, MA   02139

Library
Lamont-Doherty Geological Observatory
Colombia University
Palisades, NY   10964

Library
Serials Department
Oregon State University
Corvallis, OR   97331

Pell Marine Science Library
University of Rhode Island
Narragansett Bay Campus
Narragansett, RI   02882

Working Collection
Texas A&M University
Dept. of Oceanography
College Station, TX   77843

Library
Virginia Institute of Marine Science
Gloucester Point, VA   23062

Fisheries-Oceanography Library
151 Oceanography Teaching Bldg.
University of Washington
Seattle, WA   98195

Library
R.S.M.A.S.
University of Miami
4600 Rickenbacker Causeway
Miami, FL   33149

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. WHOI-86-26	2.	3. Recipient's Accession No.
4. Title and Subtitle Multiple Source Location Estimation Using the EM Algorithm			5. Report Date July 1986
			6.
7. Author(s) Ehud Weinstein and Meir Feder			8. Performing Organization Rept. No. WHOI-86-26
9. Performing Organization Name and Address Woods Hole Oceanographic Institution Woods Hole, Massachusetts 02543			10. Project/Task/Work Unit No.
			11. Contract(C) or Grant(G) No. (C) N00014-85-K-0272 (G)
12. Sponsoring Organization Name and Address Naval Air Systems Command			13. Type of Report & Period Covered Technical
			14.

15. Supplementary Notes

This report should be cited as: Woods Hole Oceanog. Inst. Tech. Rept. WHOI-86-26.

16. Abstract (Limit: 200 words)

We present a computationally efficient scheme for multiple source location estimation based on the EM Algorithm. The proposed scheme is optimal in the sense that it converges iteratively to the exact Maximum Likelihood estimate of all the unknown parameters simultaneously. The method can be applied to a wide range of problems arising in signal and array processing.

17. Document Analysis   a. Descriptors
1. Array Processing
2. Multiple Source Localization
3. EM Algorithm

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 27
Approved for publication; distribution unlimited.	20. Security Class (This Page)	22. Price

(See ANSI-Z39.18)   See Instructions on Reverse

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce